Analyzing the Effects of Pixel-Scale Data Fusion in Hyperspectral Image Classification Performance

Michael A. Alvarez[a,b,d], Carlos A. Theran[a,c,d], Emmanuel Arzuaga[a,b,c,d], and Heidy Sierra[a,c,d,*]

[a]Laboratory for Applied Remote Sensing, Imaging and Photonics
[b]Department of Electrical and Computer Engineering
[c]Department of Computer Science and Engineering
[d]University of Puerto Rico Mayaguez, Call Box 9000, Mayaguez, PR, 00680.

ABSTRACT

Recently, multispectral and hyperspectral data fusion models based on deep learning have been proposed to generate images with a high spatial and spectral resolution. The general objective is to obtain images that improve spatial resolution while preserving high spectral content. In this work, two deep learning data fusion techniques are characterized in terms of classification accuracy. These methods fuse a high spatial resolution multispectral image with a lower spatial resolution hyperspectral image to generate a high spatial-spectral hyperspectral image. The first model is based on a multi-scale long short-term memory (LSTM) network. The LSTM approach performs the fusion using a multiple step process that transitions from low to high spatial resolution using an intermediate step capable of reducing spatial information loss while preserving spectral content. The second fusion model is based on a convolutional neural network (CNN) data fusion approach. We present fused images using four multi-source datasets with different spatial and spectral resolutions. Both models provide fused images with increased spatial resolution from 8m to 1m. The obtained fused images using the two models are evaluated in terms of classification accuracy on several classifiers: Minimum Distance, Support Vector Machines, Class-Dependent Sparse Representation and CNN classification. The classification results show better performance in both overall and average accuracy for the images generated with the multi-scale LSTM fusion over the CNN fusion.

Keywords: Hyperspectral image, Multispectral image, Fully Connected Network, Data fusion, Machine learning learning, Decision fusion

***Heidy Sierra, heidy.sierra1@upr.edu**

1. INTRODUCTION

Given the high availability of data captured by different remote sensing instruments, the classification of multi-spectral and hyperspectral images (HSI) is an important topic in remote sensing. Different classifiers have been proposed for this task.[1] Nevertheless, the classification of HSI can become an arduous task for different reasons. For example, the availability of reference data (validation data) often implies a short or limited amount of training samples as well as redundant features. It is well understood that as the number of dimensions increases, the number of samples needed to have high classification accuracy increases as well.[2,3]

Traditional classifiers include: Minimum Distance (e.g. Euclidean, City block and Mahalanobis), Maximum Likelihood and Support Vector Machines (SVM).[3,4] Recently, class-dependent Sparse Representation Classifiers

(cdSRC)[5–7] as well as Deep Learning methods such as Convolutional Neural Networks (CNN)[8] have shown to provide high classification accuracy.

With the recent increase of sensors available, the fusion of these multiple sources of data is being adopted as a mechanism to improve the classification results.[9,10] In particular, HSI and Multispectral Images (MSI) are fusing to generate Hyperspectral Enhanced Images (HSEI), which have both a high spatial resolution and a high spectral resolution.[11,12] Recently, multispectral and hyperspectral data fusion models based on deep learning have been proposed.[13,14] These techniques generate images with high spatial and spectral resolution. However, a study regarding the effects of such fusions in terms of classification accuracy has not been presented.

In this work, two deep learning data fusion techniques to generate HSEI are characterized in terms of classification accuracy. These methods fuse a high spatial resolution multispectral image with a lower spatial resolution hyperspectral image to generate a high spatial-spectral hyperspectral image. The first model performs the HSEI data fusion based on a Long Short-Term memory (LSTM) network.[13] The second fusion model uses a convolutional neural network (CNN)[14] to generate de HSEI. For the images generated by both models, we compute the Average Accuracy (AA) and Overall Accuracy (OA) as metrics of classification performance using several supervised classifiers.

2. FRAMEWORK

As mentioned, two deep learning data fusion techniques are being employed to generate the HSEIs used in this paper. These data fusion processes attempt to improve results from a traditional pansharpening method. In particular, a component substitution method is adopted in which the network learns to enhance the spatial-resolution of an HS image by integrating the data of interest provided by a MS image. The deep learning techniques based on LSTM and CNN networks are presented in subsections 2.1 and 2.2, respectively.

We use Minimum Euclidian Distance, SVM (with the linear kernel), cdSRC (with city block distance as regularization) and CNN classifiers to measure the effects of using HSEI for supervised classification. The classification process we followed is described in details in subsection 2.3.

2.1 LSTM for spatial-resolution enhancement

The LSTM approach fuses HSI and MS using a multiple-step process. This technique does a transition from low to high spatial resolution using an intermediate step. It is capable of reducing spatial information loss while preserving spectral content.[13]

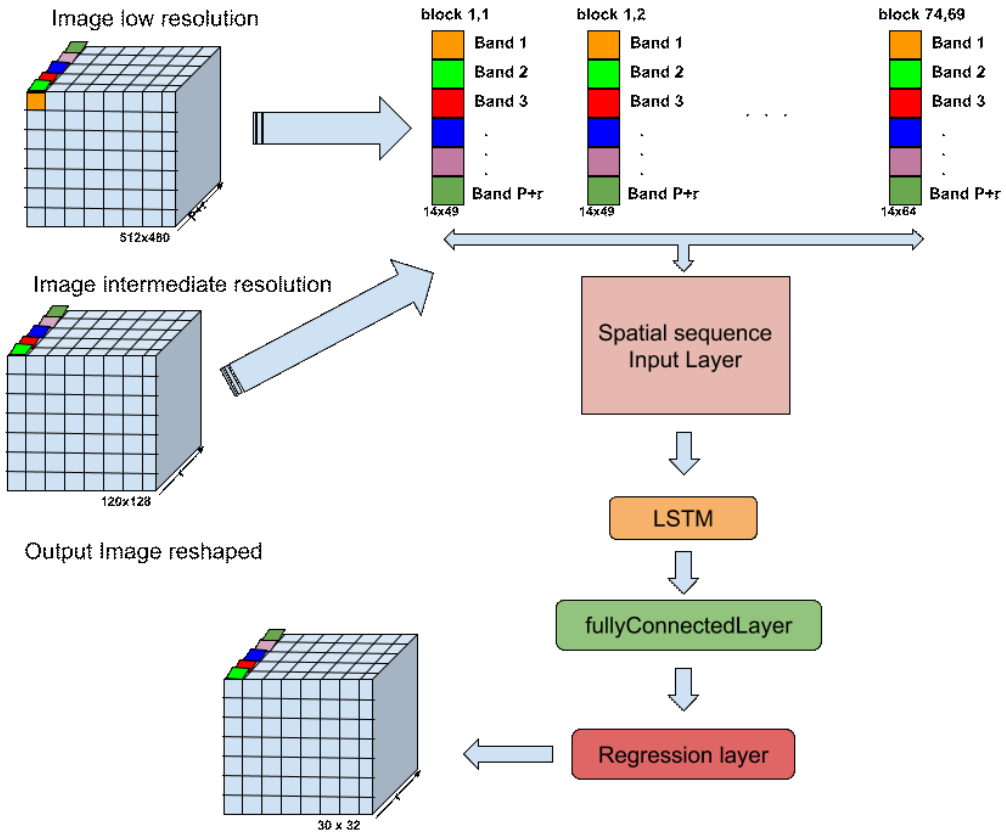


Figure 1: LSTM architecture for high spatial resolution images

This architecture is designed to retain information from low resolution to intermediate resolution in order to get to the desired high resolution. Figure 1 presents the model for high spatial resolution images based on LSTM.

2.2 CNN for spatial-resolution enhancement

The CNN network is a deep learning technique that integrates the variability of 2D shapes.[15] The CNN method to fuse MS and HSI images to obtain HSEI was proposed by Palsson.[14] The CNN network employed consists of three convolutional layers, with 32, 64, and 10 filters, respectively.

2.3 Classifiers

- Minimum Euclidian Distance

  Minimum distance classifiers belong to a family of classifiers referred to as sample classifiers.[16] The classification of each pixel depends on a group measure, more than its particular spectral response. Let S be the set of pixels-samples whose class is known. For each class compute statistical properties (as mean $\hat{x}_i$, variance σ_i), then the minimum distance decision rule consists of computing the d_i distances of each pixel to each subset sample observations, the class of the pixel corresponding to the min value calculated. The Euclidian Distance d_i is usually defined as:

$$d_i = \sqrt{\sum_{j=1}^{N}(x[j] - \hat{x}_i[j])^2} \ , \quad \forall i \in \text{classes of } S \tag{1}$$

  where $x \in \mathcal{R}^N$ is the pixel-vector, $\hat{x}_i \in \mathcal{R}^N$ is the mean value of each class.

- SVM

  The SVM classifier is a generalization of linear decision boundaries for classification. SVM computes the best hyperplane that separates all data points in two classes. Finding the best hyperplane means to identify the hyperplane with the largest margin between classes. This produces nonlinear boundaries by the construction of a linear boundary in a transformed version of the feature space.[17]

  The mathematical formulation of the two-class problem[18] begins with the hyperplane $f(x)$ defined as:

$$f(x) = x^T\beta + \beta_0 = 0 \tag{2}$$

  where $x_j, \ \beta \in \mathcal{R}^N \ , \ \beta_0 \in \mathcal{R}$.

  For each pixel x_j, their respective class is labeled as $y_j = \pm 1$. Find β and β_0 that minimize $\|\beta\|$ subjet to:

$$y_j * f(x_j) \geq 1 \quad \text{, for all data points } (x_j, y_j). \tag{3}$$

  Additionally, the values x_j at the boundary $y_j * f(x_j) = 1$ are called support vectors.

For multiclass problems, the main idea is to solve iteratively many two-class problems.[17] There are several ways to compute SMV classifier using different loss functions and suitable kernels (Gaussian, Linear, Polynomial) according to application.[19] In this work, we use a linear kernel, and *fitcecoc* function available as part of the Statistics and Machine Learning Toolbox in MATLAB R2015a.

- cdSRC

In the SRC method, each test pixel is approximated by the span of training pixels of each class,[5] i.e., let y be a test pixel, let X_j be a matrix formed by the train pixels of class j, then

$$y \approx X_j \alpha_j^T \tag{4}$$

where α_j is a vector who project the test pixel y in the span of the train pixels matrix X_j. By extending the projection over training pixels of all classes,

$$y \approx X\alpha \tag{5}$$

where $X = [X_1,\ X_2,\ \ldots,\ X_N]$ and $\alpha = [\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_N]$. The approximation (5) of y must be better than (4) because generated space by all classes contains each span of X_j. Now the solution of the (5) for each test pixel produces the classification of the data.

In this work, we use the cdSRC method using the Moore-Penrose pseudoinverse, to solve (5) (proposed by Arias, Sierra and Arzuaga[6]). This implementation reduces the time complexity and obtains a comparable classification accuracy with the state of art methods.

- CNN

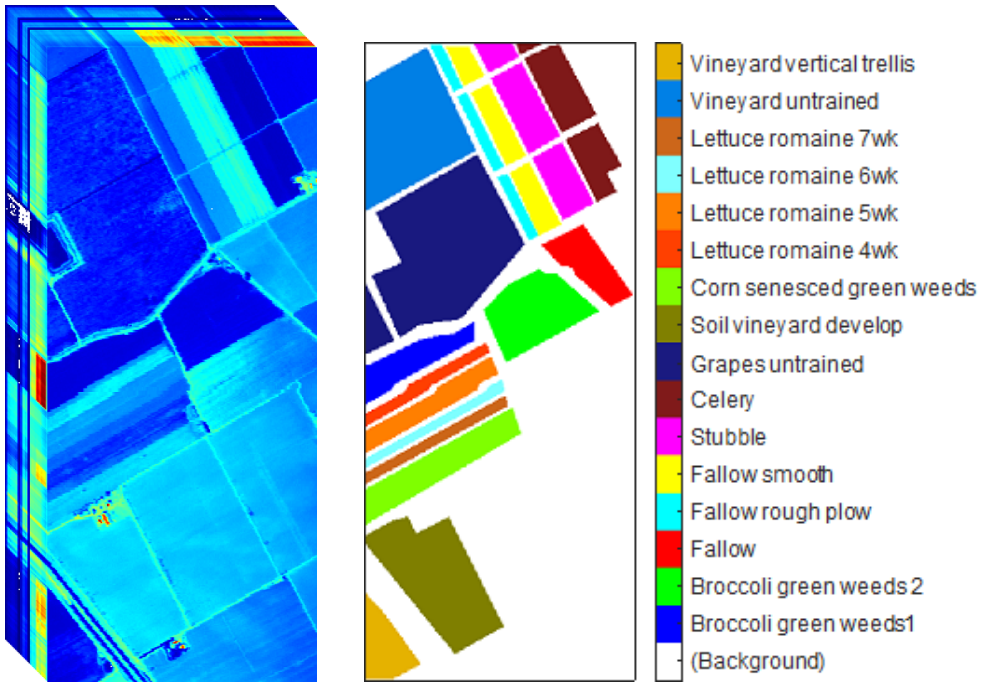
The CNN classification method, integrates the variability of 2D shapes, which means high performance in recognition problems, and an intrinsic architecture adequate to classify hyperspectral images.[20] The layers used are [conv - relu - maxpool] x 2 - [affine - relu] x 2 - affine - softmax, this configuration is state of art.[21]

We use 11 pixels as a block-size and the Deep Learning Toolbox in MATLAB R2019b.

3. EXPERIMENTAL DATA

The experimental data used in this work are four different hyperspectral images: Indian Pines, Salinas, and Enrique Reef. These images are described as follows:

- The Salinas hyperspectral image has a spatial resolution of 3.7m, 204 bands, and was collected by the AVIRIS sensor. It consists of 512×217 pixels and 16 classes, illustrated in Figure 2b.

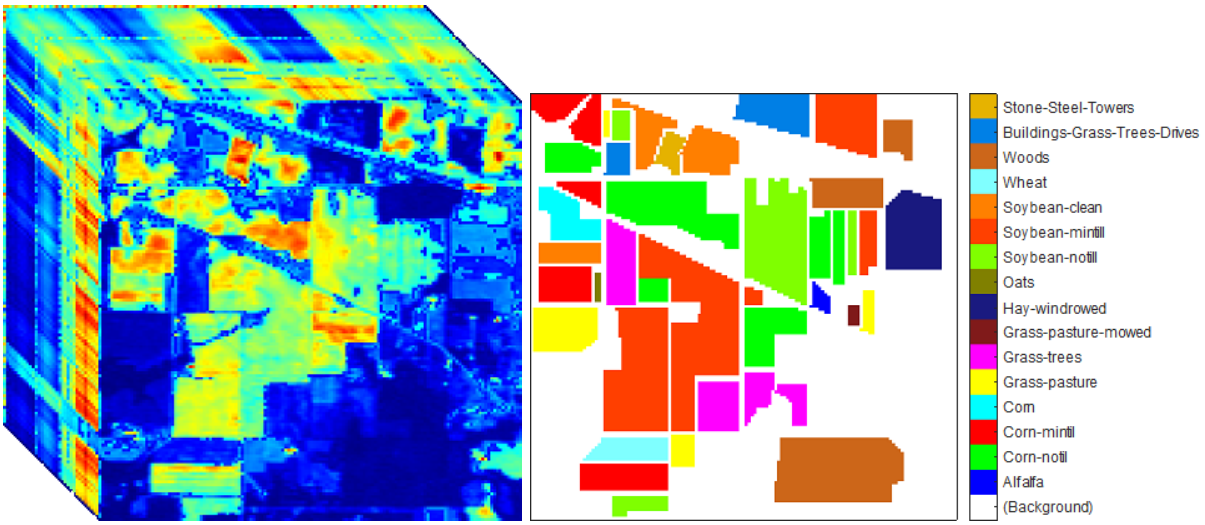


(a) (b)

Figure 2: 2a The Salinas hyperspectral dataset. 2b The Salinas classification map

- The Indian Pines hyperspectral image was gathered by the AVIRIS sensor and consists of 145×145 pixels and 224 spectral bands in the wavelength range of 400 to 2500 nm. The number of bands was reduced to 200 by removing high water absorption bands. This scene has 16 classes depicted in Figure 3b.

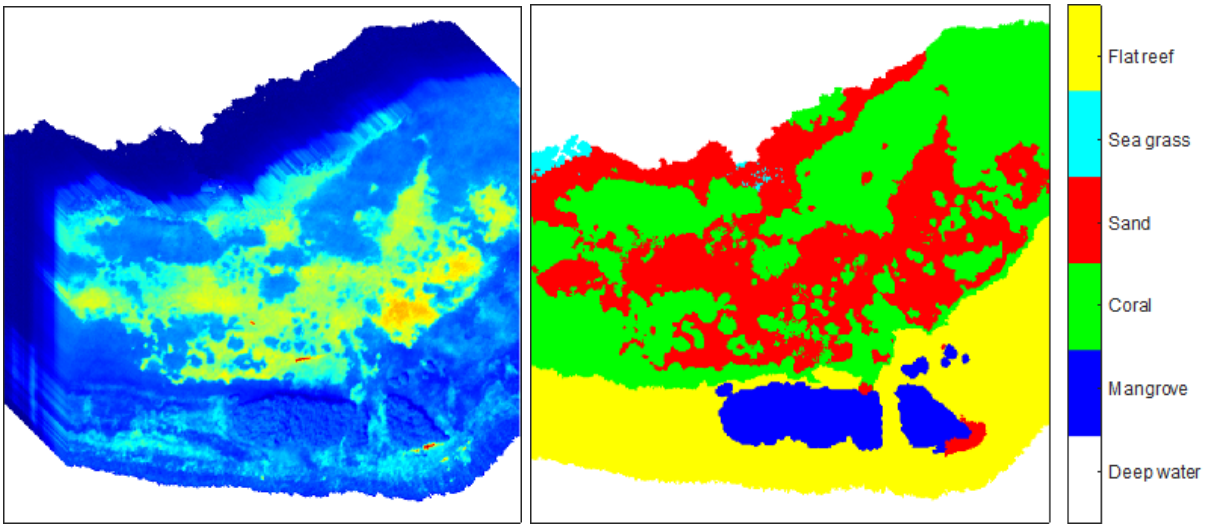
- The Enrique Reef hyperspectral image consists of 120 bands taken from the AISA Eagle sensor with a spatial resolution of 1, 2, 4, and 8m. The images were captured by the Galileo group in 2007. For our



(a) (b)

Figure 3: 3a Indian Pines HSI. 3b Indian Pines Classification Map.

tests, we used the high-resolution hyperspectral image (1m), and the low-resolution hyperspectral image (8m). There are six classes illustrated in Figure 4b.



(a) (b)

Figure 4: 4a Enrique HSI . 4b Enrique Classification Map

All of these images are used to generate simulated data. A low spatial resolution hyperspectral image is simulated by applying image decimation to the original dataset following the procedure presented in.[13,14] Likewise, a high spatial resolution multispectral image is simulated from the original data set by averaging bands from different spectral ranges: 1) Blue: 445-516nm, 2) Green: 506-595nm, 3) Red: 632-698nm and 4) NIR = 757-853nm. Both

simulated images are then fused using the CNN and LSTM techniques[13,14] to generate two HSEI. Since the Enrique Reef dataset provides real data from 1m and 8m a fourth experiment is conducted with real sensor data by using the 8m hyperspectral dataset as the low spatial resolution image and simulating the multispectral image with the 1m dataset.

4. EXPERIMENTAL PROCEDURE

To provide a performance analysis in terms of classification, we classify the HSEI generated images from the datasets presented in Section 3. The low-resolution hyperspectral image of each dataset is enhanced using the deep learning data fusion techniques, and the obtained HSEI was classified using Minimum Euclidean Distance (MD), Support Vector Machine (SVM), class-dependent Sparse Representation Classifier (cdSRC), and CNN classifiers. The MD and SMV are traditional classifiers while cdSRC and CNN are state of the art classifiers in current HSI classification literature, which should provide an excellent range of different supervised classification schemes for comparison in classification terms. Furthermore, the original dataset is also classified and compared with the classification of the HSEI cube.

5. EXPERIMENTAL RESULTS

The classificaiton experiments considered two main cases: first PCA feature extraction transformation was performed, reducing data dimensionality to 20 bands (since some sets have up to sixteen classes), and the second case used the full bands range. Additionally, we gradually change the amount of data to train each classifier: 10, 30, 50 and 75 percent of the data randomly chosen. The average and overall classification accuracy are presented for each case.

5.1 Indian Pines Hyperspectral Dataset Results

Table 1 presents the Overall and Average Accuracy, for each classifier, for the Indian Pines reference dataset. These values are used to compare how much the classification improves by using a model to improve spatial resolution. Tables 2 and 3 present the Overall and Average Accuracy, respectively, of the generated HSEI using the Indian Pines dataset reduced to 20 bands.

	Overall Accuracy				Average Acuracy			
Train / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	39.79%	61.10%	56.66%	66.84%	48.91%	50.04%	39.62%	56.81%
30 / 70	39.45%	61.98%	65.45%	73.71%	49.66%	55.37%	47.22%	74.58%
50 / 50	38.74%	62.60%	73.01%	83.66%	47.78%	54.60%	52.37%	85.63%
75 /25	39.89%	63.32%	80.64%	87.17%	49.87%	56.69%	62.78%	90.76%

Table 1: Classification results for Indian Pines reference data, reduced to 20 bands. Each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

In the same manner, tables 4 and 5 present the Overall and Average Accuracy, respectively, of the generated HSEI using the Indian Pines dataset spanning the full bands range.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	26.45%	66.39%	61.91%	53.46%	42.78%	70.31%	75.19%	70.68%
30 / 70	23.95%	68.58%	74.37%	58.15%	43.09%	71.42%	89.15%	77.23%
50 / 50	24.84%	69.57%	82.42%	64.83%	44.38%	72.05%	93.02%	86.38%
75 /25	24.45%	69.36%	88.48%	71.79%	45.16%	72.32%	96.59%	90.79%

Table 2: Experiment with Indian Pines dataset, reduced to 20 bands. The overall accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	30.77%	56.32%	38.56%	37.87%	51.30%	69.95%	57.48%	62.57%
30 / 70	26.91%	63.53%	53.64%	47.74%	52.50%	71.19%	83.62%	78.29%
50 / 50	30.04%	66.39%	62.39%	55.90%	52.31%	75.43%	84.14%	90.69%
75 /25	28.18%	67.05%	74.36%	72.84%	52.55%	76.52%	94.88%	90.75%

Table 3: Experiment with Indian Pines dataset, reduced to 20 bands. The average accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	26.65%	92.00%	86.61%	73.15%	44.29%	94.63%	93.21%	75.58%
30 / 70	24.28%	96.39%	97.00%	87.35%	44.90%	96.39%	98.38%	98.98%
50 / 50	25.06%	99.03%	98.50%	93.60%	46.05%	99.57%	99.39%	99.80%
75 /25	24.69%	97.83%	99.00%	93.95%	46.96%	99.79%	99.76%	99.95%

Table 4: Experiment with Indian Pines dataset, full bands range. The overall accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

We can observe from tables 1-5, of Indian Pines classification results, that the HSEI using both methods manage provide better classification results than using the original dataset. This results validate that the

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	31.35%	81.07%	67.62%	61.98%	54.44%	86.48%	84.04%	76.55%
30 / 70	27.82%	95.08%	85.90%	86.46%	54.42%	95.08%	97.44%	99.15%
50 / 50	30.57%	96.65%	95.68%	94.72%	53.96%	97.92%	98.86%	99.74%
75 /25	28.85%	98.26%	97.57%	95.02%	54.31%	98.70%	99.71%	99.78%

Table 5: Experiment with Indian Pines dataset, full bands range. The average accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

enhancement produced by the data fusion methods provide a higher quality image which enables a better classification of the spatial scene. We can also observe that the LSTM approach provides better classification results than using the other technique based on a CNN. This case is true even for a simple classifier such as MD, which improves between 3% and 6%, as well as a robust classifier such as cdSRC, with increases between 16% and 24%.

5.2 Salinas Hyperspectral Dataset Results

Table 6 presents the Overall and Average Accuracy, for each classifier, for the Salinas reference dataset. These values are used to compare how much the classification improves by using a model to improve spatial resolution. Tables 7 and 8 present the Overall and Average Accuracy, respectively, of the generated HSEI using the Salinas dataset reduced to 20 bands.

	Overall Accuracy				Average Acuracy			
Train / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	75.56%	91.71%	48.55%	99.21%	80.51%	95.38%	47.17%	98.77%
30 / 70	75.45%	91.98%	54.01%	99.65%	80.66%	95.90%	55.53%	99.40%
50 / 50	75.46%	92.07%	60.35%	99.90%	80.73%	96.00%	63.90%	99.83%
75 /25	75.79%	92.08%	64.41%	99.89%	80.95%	96.07%	60.62%	99.85%

Table 6: Classification results for Salinas reference data, reduced to 20 bands. Each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

We can observe that both image enhancement methods provide an HSEI capable of producing better classification results than the original dataset. Given the spectral characteristics of the Salinas set, a CNN classifier with only 10% of training data obtains an OA of 99.21%, and for the same samples in the HSEI an OA of 99.37%, indicating that the enhancement techniques improve specific details that help classifiers better separate between classes. For this case in particular both methods provide similar results execpt for the cdSRC classifier in which

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	70.06%	93.17%	79.18%	99.22%	75.94%	93.77%	78.46%	99.35%
30 / 70	70.44%	93.52%	91.95%	99.69%	75.96%	93.81%	67.95%	99.83%
50 / 50	70.65%	93.67%	92.17%	99.93%	76.01%	94.07%	74.99%	99.88%
75 /25	70.69%	93.79%	90.01%	99.97%	75.97%	94.12%	80.74%	99.97%

Table 7: Experiment with Salinas dataset, reduced to 20 bands. The overall accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	73.74%	95.84%	83.12%	98.66%	79.08%	96.69%	80.54%	98.96%
30 / 70	74.21%	96.49%	87.43%	99.32%	79.29%	97.19%	81.46%	99.62%
50 / 50	74.29%	96.66%	93.74%	99.88%	79.29%	97.35%	82.99%	99.81%
75 /25	74.34%	96.81%	93.61%	99.95%	79.33%	97.41%	83.84%	99.94%

Table 8: Experiment with Salinas dataset, reduced to 20 bands. The average accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

the CNN method yielded higher accuracy. In the case of the cdSRC classifier in particular, the original data has an OA of 48%, which becomes 79%, due to the positive impact that the enhanced spatial details has on the classification using only 10% of training data. The behavior is consistent using more data for training in the presented cases, as can be seen in tables 6,7,8. The case of the generated data with full bands shows similar behavior as the one presented, the tables are not included because of space limitations in the manuscript.

	Overall Accuracy				Average Acuracy			
Train / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	56.61%	79.79%	84.35%	88.76%	65.20%	65.62%	77.15%	71.92%
30 / 70	56.37%	79.97%	88.50%	88.61%	65.11%	68.78%	84.14%	72.34%
50 / 50	56.67%	79.94%	91.48%	89.49%	65.21%	66.44%	87.34%	73.52%
75 /25	56.41%	79.75%	94.72%	89.62%	65.98%	65.74%	91.53%	73.56%

Table 9: Classification results for Enrique reference data, reduced to 20 bands. Each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

5.3 Enrique Reef Hyperspectral Dataset Results

Table 9 presents the Overall and Average Accuracy, for each classifier, for the Enrique reference dataset. Tables 10 and 11 present the Overall and Average Accuracy, respectively, of enhanced Enrique Reef dataset with simulated 8m HSI reduced to 20 bands. The classification results are similar when the classifiers take a significative amount of data training, obtaining the better result by LSTM in order of 2%, when the data training is reduced (like

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	65.55%	65.07%	80.79%	63.65%	65.91%	75.85%	82.86%	72.97%
30 / 70	65.52%	71.97%	85.97%	73.05%	65.86%	74.48%	88.55%	73.19%
50 / 50	65.59%	72.05%	89.23%	74.01%	65.94%	75.67%	91.19%	74.52%
75 /25	65.62%	71.14%	92.56%	74.27%	66.04%	66.20%	93.98%	74.91%

Table 10: Experiment with Enrique Reef dataset with simulated 8m HSI, reduced to 20 bands. The overall accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	57.09%	80.91%	85.66%	77.39%	57.36%	81.99%	87.16%	89.53%
30 / 70	56.86%	81.11%	89.96%	89.31%	57.14%	82.17%	91.40%	89.45%
50 / 50	57.04%	81.07%	92.89%	89.76%	57.34%	82.14%	93.91%	90.44%
75 /25	56.81%	81.09%	95.78%	89.95%	57.13%	81.95%	96.24%	90.73%

Table 11: Experiment with Enrique Reef dataset with simulated 8m HSI, reduced to 20 bands. The average accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

10/90 case), the accuracy increases up to 12%.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	52.05%	73.26%	86.61%	77.39%	57.25%	82.35%	87.06%	89.72%
30 / 70	52.06%	73.35%	91.95%	78.37%	57.04%	82.45%	91.48%	89.04%
50 / 50	52.19%	73.37%	94.50%	79.03%	57.27%	82.42%	93.95%	90.27%
75 /25	52.21%	73.33%	96.79%	79.66%	57.01%	82.23%	96.31%	90.64%

Table 12: Experiment with Enrique Reef dataset, reduced to 20 bands. The overall accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

Enhancement Method	CNN 8-1				LSTM			
Training / Testing	ED	SVM	cdSRC	CNN	ED	SVM	cdSRC	CNN
10 / 90	61.23%	59.02%	78.20%	63.65%	65.94%	76.32%	82.62%	72.99%
30 / 70	61.55%	59.18%	88.76%	64.20%	65.90%	74.99%	88.64%	72.71%
50 / 50	61.40%	59.24%	92.35%	66.60%	65.97%	75.82%	91.00%	74.48%
75 /25	61.73%	59.21%	95.19%	65.89%	66.10%	66.69%	93.86%	74.75%

Table 13: Experiment with Enrique Reef dataset, reduced to 20 bands. The average accuracy is shown for each method of spatial enhancement resolution: CNN and LSTM, each row represents an amount of training data, i.e., 10, 30, 50, and 75 percent, respectively, and each column represents a classifier, i.e., MD, SVM, cdSRC and CNN classifiers, respectively.

5.4 Enrique Reef Hyperspectral Dataset using Real Data

As was mentioned in Section 3, the Enrique Reef dataset includes several images of the same spatial scene with different spatial resolutions. In this last experiment we take advantage of the existence of this real dataset to produced real HSEI using both techniques (CNN and LSTM). The data fusion techniques use the 8m data as the low spatial resolution and the high spatial resolution multispectral image from the 1m spatial resolution dataset, using the same reference data mentioned in subsection 5.3.

Table 9 presents the Overall and Average Accuracy, for each classifier, for the Enrique reference dataset. These values are used to compare how much the classification improves by using a model to improve spatial resolution. Tables 12 and 13 present the Overall and Average Accuracy, respectively, of the generated HSEI using the Enrique Reef dataset reduced to 20 bands.

By comparing the results in Tables 9-13, the classification accuracy improvement is not as dramatic as the previous datasets. The CNN generated HSEI in particular does not provide an HSEI capable of improving classification accuracy. At the same time, the LSTM method allows for a better classification than the original dataset but with moderate improvements. One possible explanation to this behavior may be related with the fact that the spatial resolution of this dataset (1m, 8m) is considerably higher than the previous two datasets and is capable of providing sufficient level of detail to accurately identify the spatial scene. In this sense, both methods were not capable of identifying further spatial enhancements that provide a higher classification accuracy, with the CNN technique actually providing results that reduce classification accuracy.

6. CONCLUSIONS AND FUTURE WORK

This paper presents the impact on supervised classification performance of two deep learning based data fusion methods. We have numerically proven that the LSTM is a useful architecture to generate HSEI, compared with other architectures in literature, showing better results in classification accuracy. We also show that the LSTM cubes are better to classify than the reference HS images. These results are shows a promising use of the LSTM architecture to integrate multi sensor data to improve classification. In a sequence paper, we show a decision fusion technique that further improves overall classification by combining the multiple classifier results, obtaining

a higher accuracy per class. In future work, we propose a deep learning technique that performs image fusion to enhance spatial-spectral resolution that integrates classification results in the network training.

ACKNOWLEDGMENTS

We thank Dr. Fernando Arias for sharing the cdSRC code, and for useful discussions about classification problems in remote sensing. We also thank L3Harris Technologies Space and Airborne Systems Division in Rochester NY for their support with the access to Deep Learning Infrastructure to conduct our experiments. This material is based upon work supported by the National Science Foundation under Grant No. OAC-1750970.

REFERENCES

[1] R. G. Congalton *et al.*, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote sensing of environment* **37**(1), 35–46 (1991).

[2] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press (1990).

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley & Sons (2012).

[4] E. Carrizosa and D. R. Morales, "Supervised classification and mathematical optimization," *Computers & Operations Research* **40**(1), 150–165 (2013).

[5] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing* **53**(5), 2683–2695 (2014).

[6] F. X. Arias, H. Sierra, and E. Arzuaga, "Improving execution time for supervised sparse representation classification of hyperspectral images using the moore–penrose pseudoinverse," *Journal of Applied Remote Sensing* **13**(2), 026512 (2019).

[7] F. X. Arias, H. Sierra, L. O. Jimenez-Rodriguez, *et al.*, "Supervised sparse-representation classification on hyperspectral images using the city-block distance to improve performance," (2017).

[8] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine* **4**(2), 22–40 (2016).

[9] D. Amarsaikhan, M. Saandar, M. Ganzorig, *et al.*, "Comparison of multisource image fusion methods and land cover classification," *International Journal of Remote Sensing* **33**(8), 2532–2550 (2012).

[10] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Processing* **92**(9), 2082–2096 (2012).

[11] P. Ghamisi, B. Rasti, N. Yokoya, *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geoscience and Remote Sensing Magazine* **7**, 6–39 (2019).

[12] J. Zhang, "Multi-source remote sensing data fusion: status and trends," *International Journal of Image and Data Fusion* **1**(1), 5–24 (2010).

[13] C. Theran, M. Álvarez, E. Arzuaga, *et al.*, "A pixel level scaled fusion model to provided high spatial-spectral resolution for satellite images using LSTM netwroks," *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* , 1–5 (2019).

[14] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters* **14**, 639–643 (2017).

[15] Y. LeCun, L. Bottou, Y. Bengio, *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* **86**(11), 2278–2324 (1998).

[16] A. Wacker and D. Landgrebe, "Minimum distance classification in remote sensing," *LARS Technical Reports* , 25 (1972).

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media (2009).

[18] N. Cristianini, J. Shawe-Taylor, *et al.*, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press (2000).

[19] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing* **66**(3), 247–259 (2011).

[20] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing* **219**, 88–98 (2017).

[21] A. Santara, K. Mani, P. Hatwar, *et al.*, "Bass net: Band-adaptive spectral-spatial feature learning neural

network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing* **55**(9), 5293–5301 (2017).